**Linear Regression Assignment**

1.**Problem Statement: Calories\_consumed-> predict weight gained using calories consumed.**

1.Perform EDA-1st Moment,2nd Moment,3rd Moment & 4th Moment.

2. Scatter Diagram-Direction, Strength & Linear.

3. Correlation Coefficient Value =r (-1 to +1) thumb rule if r >0.85 then it is strong. r=0.94 it is strong.

4. Coefficient of determination is R2(0 to 1) thumb rule if R2>0.8 then it is strong. R2=0.8968 this is also strong.

I have done the analysis with R studio as follows:

R-script

# Load calories\_consumed.csv dataset

library(readr)

calories\_consumed <- read\_csv("Data Science Assignments/Linear Regression/Dataset/calories\_consumed.csv")

View(calories\_consumed)

attach(calories\_consumed)

# Exploratory data analysis

summary(calories\_consumed)

#Scatter plot

plot(calories\_consumed$`Weight gained (grams)`, calories\_consumed$`Calories Consumed`) # plot(X,Y)

#Correlation Coefficient (r)

cor(`Weight gained (grams)`, `Calories Consumed`) # cor(X,Y)

# Simple Linear Regression model

reg <- lm(calories\_consumed$`Calories Consumed` ~ calories\_consumed$`Weight gained (grams)`) # lm(Y ~ X)

summary(reg)

confint(reg,level=0.95)

predict(reg,interval="predict")

Console Window

|  |
| --- |
| > # Load calories\_consumed.csv dataset  > library(readr)  > calories\_consumed <- read\_csv("Data Science Assignments/Linear Regression/Dataset/calories\_consumed.csv")  Parsed with column specification:  cols(  `Weight gained (grams)` = col\_double(),  `Calories Consumed` = col\_double()  )  > View(calories\_consumed)  > attach(calories\_consumed)  The following objects are masked from calories\_consumed (pos = 3):  Calories Consumed, Weight gained (grams)  > # Exploratory data analysis  > summary(calories\_consumed)  Weight gained (grams) Calories Consumed  Min. : 62.0 Min. :1400  1st Qu.: 114.5 1st Qu.:1728  Median : 200.0 Median :2250  Mean : 357.7 Mean :2341  3rd Qu.: 537.5 3rd Qu.:2775  Max. :1100.0 Max. :3900  > #Scatter plot  > plot(calories\_consumed$`Weight gained (grams)`, calories\_consumed$`Calories Consumed`) # plot(X,Y)  > #Correlation Coefficient (r)  > cor(`Weight gained (grams)`, `Calories Consumed`) # cor(X,Y)  [1] 0.946991  > # Simple Linear Regression model  > reg <- lm(calories\_consumed$`Calories Consumed` ~calories\_consumed$`Weight gained (grams)`) # lm(Y ~ X)  > summary(reg)  Call:  lm(formula = calories\_consumed$`Calories Consumed` ~ calories\_consumed$`Weight gained (grams)`)  Residuals:  Min 1Q Median 3Q Max  -450.41 -115.03 -41.46 194.55 375.75  Coefficients:Y=B0+B1X  **1.Point Estimation is CC=1577.201+2.134(Weight gained)**  Estimate Std. Error t value Pr(>|t|)  (Intercept) 1577.201 100.541 15.69 2.33e-09 \*\*\*  calories\_consumed$`Weight gained (grams)` 2.134 0.209 10.21 2.86e-07 \*\*\*  ---  Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1  Residual standard error: 251.5 on 12 degrees of freedom  **Multiple R-squared: 0.8968,** Adjusted R-squared: 0.8882  F-statistic: 104.3 on 1 and 12 DF, p-value: 2.856e-07  > confint(reg,level=0.95)  **2.Lower Prediction Equation is CC=1358.141455+1.678994(Weight gained)**  **3. Upper Prediction Equation is CC=1796.259949+2.589852(Weight gained)**  2.5 % 97.5 %  (Intercept) 1358.141455 1796.259949  calories\_consumed$`Weight gained (grams)` 1.678994 2.589852  > predict(reg,interval="predict")  fit lwr upr  1 1807.718 1229.249 2386.187  2 2004.085 1432.376 2575.795  3 3498.181 2879.564 4116.799  4 2004.085 1432.376 2575.795  5 2217.528 1649.740 2785.316  6 1811.987 1233.697 2390.278  7 1850.407 1273.660 2427.154  8 1709.535 1126.585 2292.485  9 2857.854 2280.041 3435.668  10 3925.066 3264.781 4585.351  11 1790.643 1211.447 2369.839  12 1897.364 1322.350 2472.378  13 2324.249 1757.059 2891.439  14 3071.297 2483.085 3659.509  Warning message:  In predict.lm(reg, interval = "predict") :  predictions on current data refer to \_future\_ responses  Scatter Diagram     1. Direction-Moderate Positive Correlation 2. Strength of Correlation -Strong 3. Linear 4. Correlation Coefficient value is r=r>0.85 it is a good relationship.   I got r>0.89-It is a good relationship   1. Coefficient of Determination is R2>0.8-is Strong   I got R2>0.9 -It is a strong   1. Y=B0+B1X    1. Point Estimation is CC=1577.201+2.134(Weight gained)    2. Lower Prediction Equation is CC=1358.141455+1.678994(Weight gained.    3. Upper Prediction Equation is CC=1796.259949+2.589852(Weight gained)   p-value: 2.856e-07  **Analysis: Calories\_consumed-> predict weight gained using calories consumed. Prediction model is good** |
|  |
| |  | | --- | |  | |